Sequence Based Indexing and Retrieval Method for Text Documents

Abstract of the Disclosure

A sequence based indexing and retrieval method for a collection of text documents includes the steps of generating a query token sequence from a query; generating at least a representative token sequence from each of the documents that contain at least one token of the query token sequence; measuring a similarity between each of the representative token sequences and the query token sequence; and retrieving the text document in responsive to the similarity of the representative query token sequence with respect to the query token sequence. The similarity measurement is preformed by determining a token appearance score, a token order score, and a token consecutiveness score of the representative token sequence with respect to the query token sequence, so as to illustrate the similarity between the representative token sequence and the query token sequence for precisely and effectively retrieving the text document.

5

10

15